Annotation of the *Propionibacterium acnes 266 CP002409* Genome at Locus Tags PAZ_c00010 to PAZ_c00060)

Sharon Luo, Sherrie Wang, Zayn Jaffri, Evan Ruttenberg, and Kathleen Kramer
Williamsville East High School and Western New York Genetics in Research Partnership

University at Buffalo

Abstract

A group of four genes from the microorganism *Propionibacterium acnes 266 CP002409* (PAZ_c00010, PAZ_c00020, PAZ_c00040, PAZ_c00060) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, potential alternative open reading frames, enzymatic function, presence or absence of gene duplication and degradation, the possibility of horizontal gene transfer, and the production of an RNA product. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the genes in the group. As such, the genes appear to be correctly annotated by in the r database.

Introduction

Propionibacterium acnes 266 CP002409 is a gram positive cocoid bacterium that is strictly aerobic, non-motile, non-encapsulated, and non-endospore forming. It is mostly found in tetrad formation and requires methionine for bacterial growth. *Propionibacterium acnes*' genome codes for a wide variety of metabolic products and has the key requirements to carry out oxidative phosphorylation, the Krebs cycle, Embden-Meyerhof pathway, and the pentose phosphate pathway. Under in vitro anaerobic conditions, *Propionibacterium acnes* can grow permissively on media such as glucose, glycerol, ribose, fructose, mannose and N-acetylglucosamine. In vivo, the bacteria produces various lipases to digest excess skin oil and sebum in the pilosebaceous units (Smith).

Propionibacterium acnes is involved in the pathogenesis of acne and uses sebum as an energy source, which is elevated in people with acne. The bacteria produces specialized enzymes that help them digest the fatty acids and triglycerides that are abundant in sebum. Fatty acids and triglycerides are fermented, and short chains of fatty acids and propionic acid are released as metabolic byproducts. *Propionibacterium acnes* may also be linked to the stimulation of sebaceous glands, thereby producing additional sebum and worsening acne conditions. However, for the most part, damage caused by acne is due to the inflammation caused by the immune response to the presence of *Propionibacterum acnes* (Science of Acne).

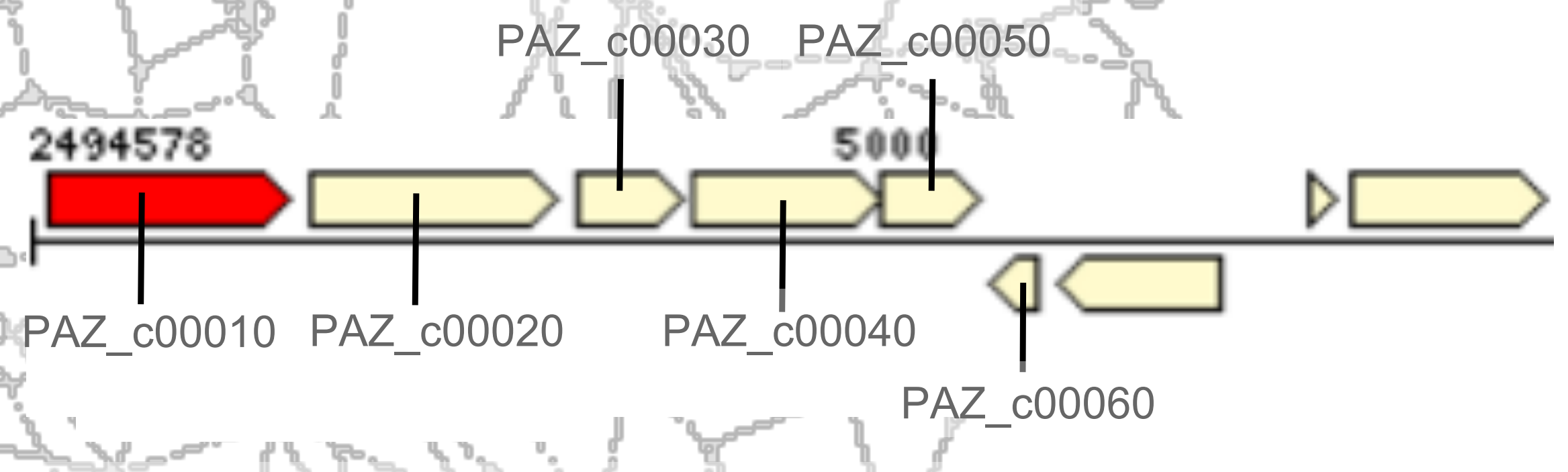


Figure 1. The locus tags and relative position of the genes under investigation in this research.

Methods

Modules of the GENI-ACT website (http://www.geni-act.org/) were used to complete *Propionibacterium acnes 266 CP002409* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 3- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 4- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate ORF Search	Has the amino acid sequence of my protein been called correctly by the computer?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

PAZ_c00010:
The initial proposed product of this gene by GENI-ACT was a 50s ribosomal protein L34. This gene product proposal was supported by the top BLAST hits, species conservation observed by its WebLogo, its cellular localization data, and its structure-based evidence. The top BLAST hits showed that the gene was present in organisms Propionibacteriaceae and *Tessaracoccus massiliensis*. Its Weblogo consists of wide, large letters throughout, indicating that the gene is well conserved throughout the species. The results of TMHMM also predicted that the protein has no transmembrane helices and most likely lies within the cell in the cytoplasm. This prediction is also supported by the predictions of PSORT-B with a large cytoplasmic score of 9.67. Finally, according to MetaCyc, the protein may be a component of the chloroplast ribosome due to its proposed pathway for cellulose and hemicellulose degradation. As such, the proposed annotation is a 50s ribosomal protein L34.

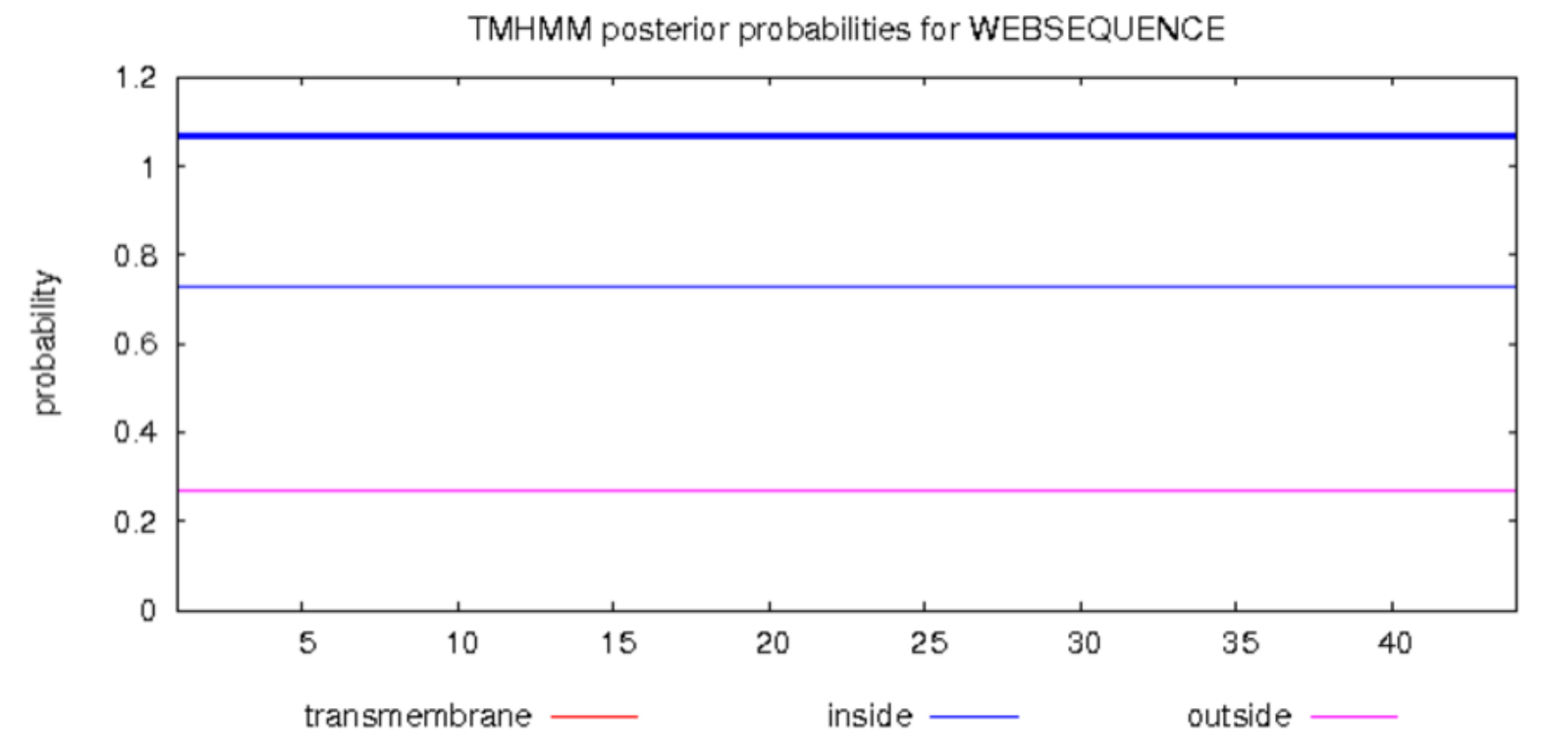
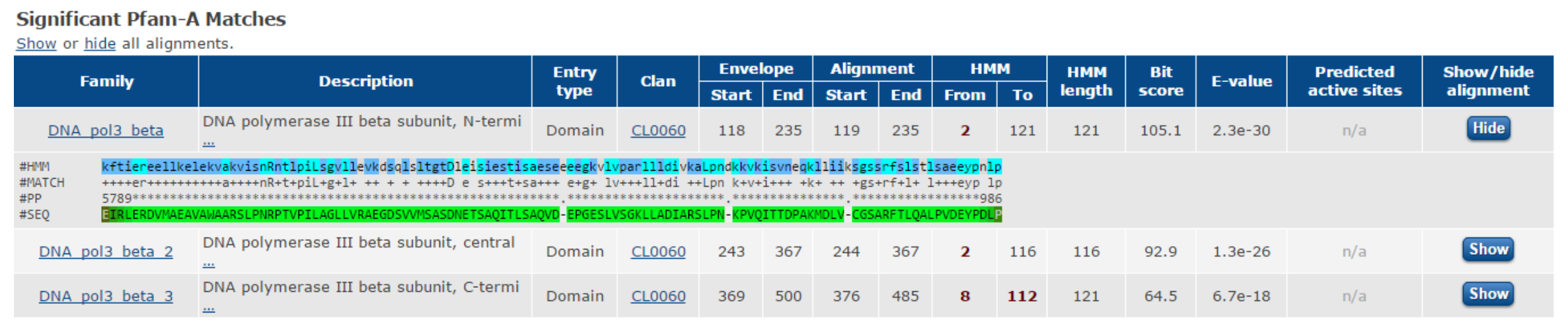


Figure 2. The thick blue line, which represents amino acids in the gene sequence predicted to be inside the cell in the cytoplasm, has a probability of approximately 0.7. Thus, it is likely for the protein to be located within the cytoplasm.

PAZ_c00020:
The initial proposed product of this gene was a DNA polymerase III beta subunit. This was supported by the top BLAST hits for the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. The top BLAST hits for this amino acid sequence were from similar species: *Propionibacterium acnes KPA171202* and *Propionibacterium acnes HL096PA1*. The hits were both DNA polymerase III beta subunits. Using Phobius, the protein was identified as non-cytoplasmic. According to KEGG, the gene codes for the enzyme DNA polymerase III, that is primarily involved with bacterial DNA replication. Significant Pfam-A matches show that the protein is part of the DNA replication pathway, as it is included in the DNA clamp clan, identified as CL0060 (Figure 3). As a result, the proposed annotation is a DNA polymerase III beta subunit.



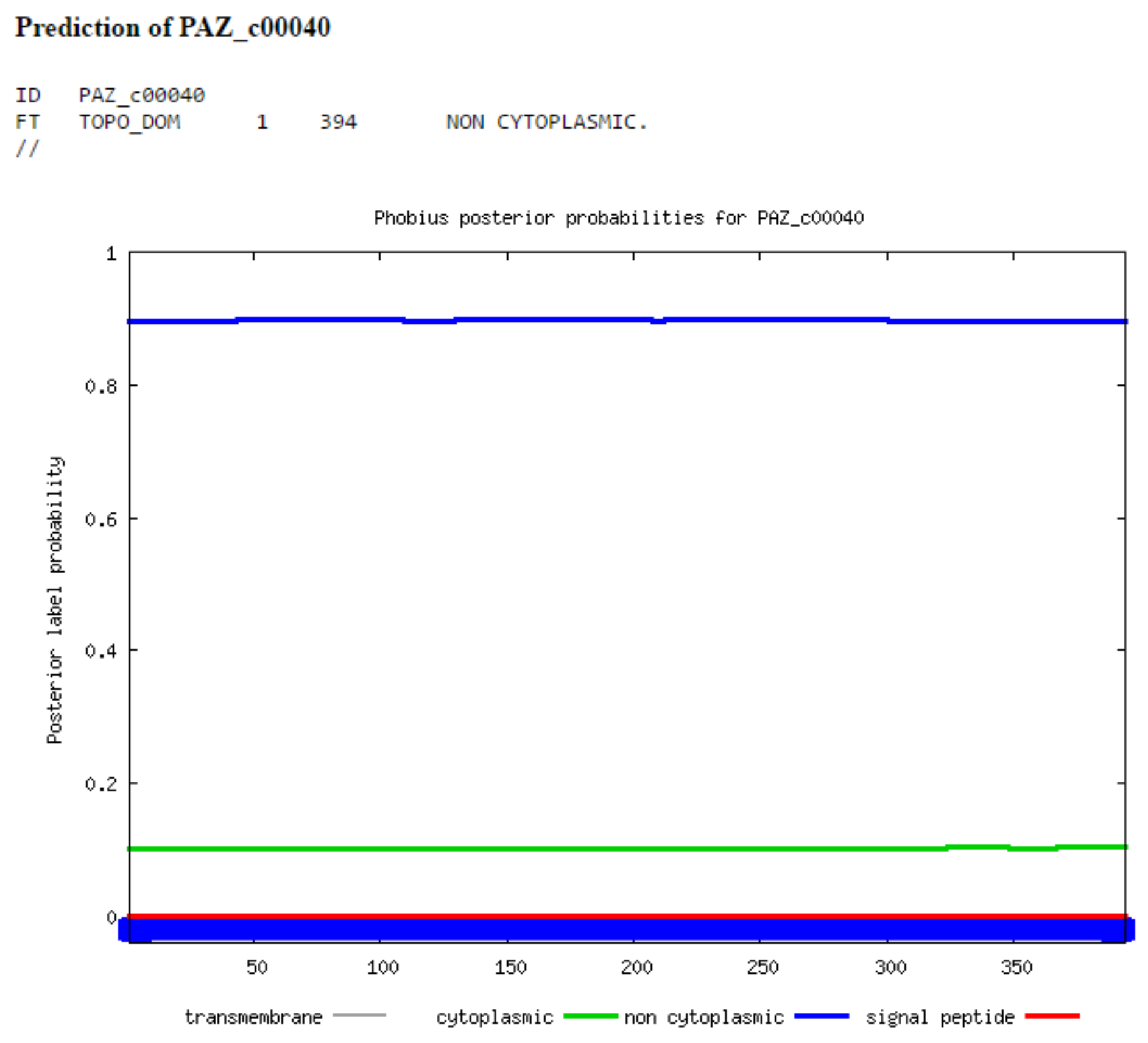
Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope Start	Envelope End	Alignment Start	Alignment End	HMM From	HMM To	HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
DNA_pol3_beta	DNA polymerase III beta subunit, N-termi ...	Domain	CL0060	118	235	119	235	2	121	121	105.1	2.3e-30	n/a	Hide
DNA_pol3_beta_2	DNA polymerase III beta subunit, central ...	Domain	CL0060	243	367	244	367	2	116	116	92.9	1.3e-26	n/a	Show
DNA_pol3_beta_3	DNA polymerase III beta subunit, C-termi ...	Domain	CL0060	369	500	376	485	8	112	121	64.5	6.7e-18	n/a	Show

```
#HMM    kftiereellkelekvakvisnRntlpiLsgvllevkdsqlsltgtDleisiestisaeseeeegkvlvparlllldivkaLpndkkvkisvneqklliiksgssrfslstlsaeeypnlp
#MATCH  ++++er++++++++++a+++++nR+t+piL+g+l+ ++ + + ++++D e s+++t+sa+++ e+g+ lv+++ll+di ++Lpn k+v+i+++ +k+ ++ +gs+rf+l+ l+++eyp lp
#PP     5789*************************************************************.*********************.**************.*****************986
#SEQ    EIRLERDVMAEAVAWAARSLPNRPTVPILAGLLVRAEGDSVVMSASDNETSAQITLSAQVD-EPGESLVSGKLLADIARSLPN-KPVQITTDPAKMDLV-CGSARFTLQALPVDEYPDLP
```

Figure 3. The top significant Pfam-A match for gene PAZ_c00020 shows that it is part of the clan CL0060, also called the DNA clamp clan. It can be concluded that PAZ_c00020 is part of the DNA replication pathway.



Prediction of PAZ_c00040

```
ID   PAZ_c00040
FT   TOPO_DOM      1    394       NON CYTOPLASMIC.
//
```

Figure 4. The blue line, which represents amino acids in the gene sequence predicted to be non-cytoplasmic, has a probability of approximately 0.9. Therefore, the PAZ_c00040 gene most likely codes for a protein not in the cytoplasm of a cell.

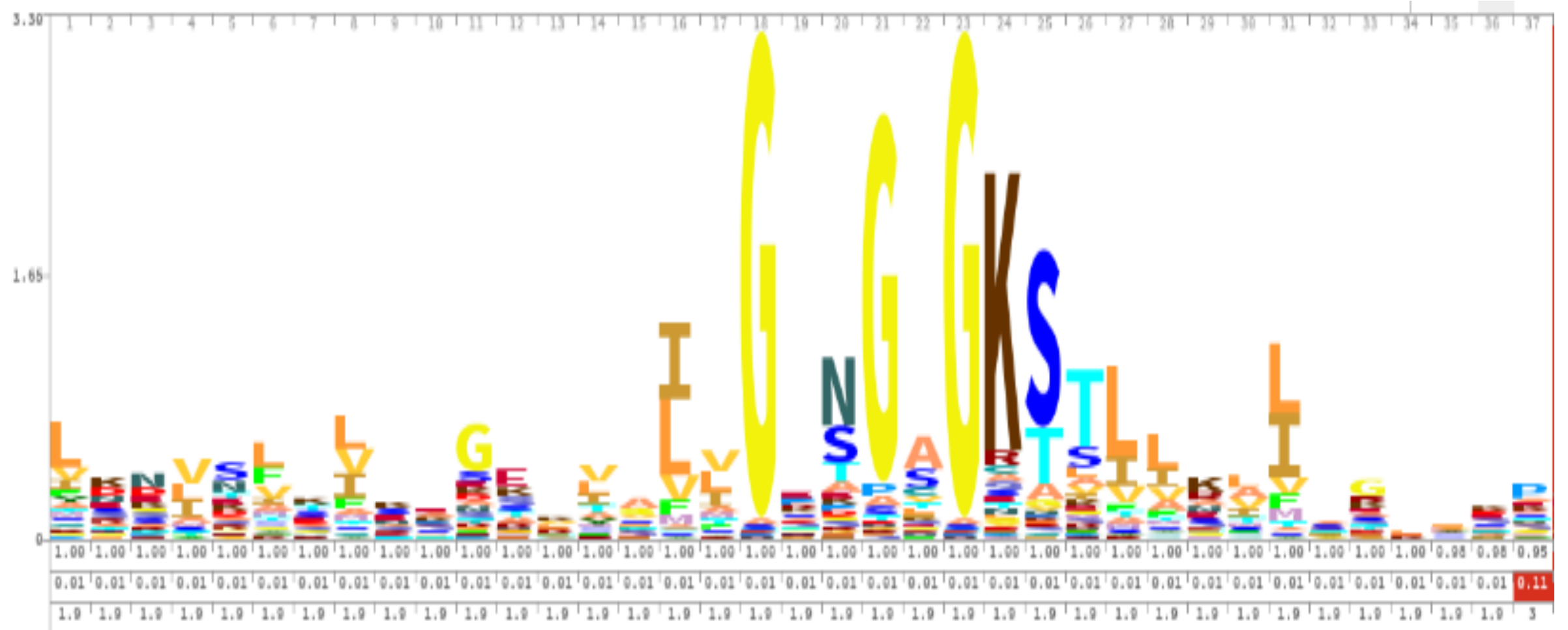


Figure 5. Denoting the relative frequency of each character, the consensus logo depicts the degree of conservation of each position using the height of the consensus character at that position.

PAZ_c00040:
The initial proposed product of this gene was a DNA replication and repair protein. This was supported by the cellular location and enzymatic function of the amino acid sequence. The protein was identified as non-cytoplasmic using Phobius (Figure 4). This gene product proposal is also supported by the identification of PAZ_c00020 as a DNA polymerase III beta subunit. As such, the proposed annotation is a DNA replication and repair protein.

PAZ_c00060:
The initial proposed product of this gene was an ABC transporter ATP-binding protein. This can be found in Propionibacteriaceae and *Cutibacterium acnes*. This was supported by the top BLAST hits for the amino acid sequence, the presence of functional domains within the amino acid sequence, and the enzymatic function of the amino acid sequence. This gene product proposal is also supported by the identification its Pfam number which is PF00005 in the clan number CL0023. The Pfam name is ABC transporter which is part of the ATP-binding protein, and as such, the proposed annotation is an ABC transporter ATP-binding protein.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group. As such, the genes appear to be correctly annotated by the computer database.

Gene Locus	GENI-ACT Gene Products	Gene Annotation
PAZ_c00010	50s ribosomal protein L34	50s ribosomal protein L34
PAZ_c00020	DNA polymerase III beta subunit	DNA polymerase III beta subunit
PAZ_c00040	DNA replication and repair protein	DNA replication and repair protein
PAZ_c00060	ABC transporter ATP-binding protein	ABC transporter ATP-binding protein

References

Smith, Christopher B. "Propionibacterium Acnes." MicrobeWiki. Kenyon College, 1 July 2011. Web. 18 May 2017.

"What Is Propionibacterium Acnes?" Science of Acne. Science of Acne, n.d. Web. 18 May 2017.

Acknowledgments

The Williamsville East Science Department would like to thank Dr. Stephen Koury for assisting with the formation of this presentation.

Supported by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) Award - 1311902

www.buffalo.edu